EuroVis Workshop on Visual Analytics (2016)
N. Andrienko and M. Sedlmair (Editors)

A Visual Analytics System for Mobile Telecommunication Marketing Analysis

M. Angelini[1] R. Corriero[2] F. Franceschi[2] M. Geymonat[2] M. Mirabelli[2] C. Remondino[3] G. Santucci[1] B. Stabellini[3]

[1]University of Rome La Sapienza, Italy
[2]Telecom Italia, Italy
[3]Polythecnic of Turin, Italy

Abstract

The Mobile Operators European market is likely one of the most competitive arenas, and even big telecommunication companies are continuously tuning their marketing strategies to contrast the aggressive campaigns of old and new competitors. This paper presents a visual analytics solution developed for supporting one of the decision making processes of TIM (Telecom Italia Group), the biggest Italian provider of telecommunications services with over 30M active mobile subscribers (at September 2014). The proposed system uses information coming from both public open data and internal TIM data, mapping them on the Italian hierarchy of 20 regions and 110 provinces. The system has been designed together with TIM analysts and allows for selecting an optimal set of provinces on which to focus marketing campaigns, according to the campaign goals, the available market, the local economy, the actual traffic, and the actual TIM penetration.

Keywords: Visual Analytics, Information Visualization, Marketing, Telecommunication

1. Introduction

The increasing competition in the market of mobile phone providers pushes telecommunication companies to continuously tune their marketing strategies to contrast the aggressive campaigns of their competitors. Such a process is a complex one and telecommunication companies rely on software solutions able to analyze data, providing suggestions and hints on the best marketing strategies. This paper presents a visual analytics solution developed for supporting one of the decision making processes of TIM. The proposed system has been developed collaboratively within a year by University of Rome, Polythecnic of Turin, and TIM, analyzing large sets of data traffic and TIM users distribution across Italy. Such data has been integrated with information coming from open data and mapped on the Italian hierarchy of 20 regions and 110 provinces. The system allows for interactively selecting an optimal set of provinces according to an objective function that is computationally intractable. In order to have an interactive system, the analytical solution allows for visually focusing on a subset of provinces (40/50) and it has been developed following an *Incremental Visualization* approach (see,e.g, [SASS15]), allowing for a real time interaction with the system and leaving the full optimization phase to a post processing step. The system is now running at TIM premises. The paper is structured as follows: Section 2 discusses related proposals, Section 3 describes the TIM scenario, Section 4 presents the implemented prototype, and Section 5 concludes the paper.

2. Related Work

The problem of coping with economic decision support and market exploitation is a well-known problem for both visual analytics [KKEM10] [KBB*10] and business intelligence fields [LT10]. Different contributions exist in the application of visual analytics to economic decision support, at different levels of detail: in [SME08] a visual analytics solution is applied to two well known decision-making problems, using simple visualizations; [SLFE11] presents a broader experimental approach to portfolio decision making using a comprehensive visual analytics dashboard. [RSE09] follows the same approach, but narrowing the analysis on personal budget basis. Most of these works cope with financial environments and do not take into consideration the problem of optimally allocating a budget in order to maximize a set of features (revenues, profits,etc.). Another point of difference with this work is that none of them is instantiated to the mobile telecommunications field nor presents a real stakeholder. Various visual approaches are considered, ranging

DOI: 10.2312/eurova.20161117

from classic charts to more complex visual representations: in particular, [ZNK08] uses pixel based visualizations in order to solve scalability problems in comparing long-term investments. Of particular interest is the work in [KMJE12]: it presents a comprehensive visual analytics environment for analysing the market. Differences with our approach are that our solution is based on competitive intelligence and shared data, a case not common in the mobile telecommunication market, given the confidentiality of these data. While designing the system, particular attention has been put on the parallel coordinates, relying on the work in [HW13] and [KBH06]; moreover, in order to insure interactivity, the developed system presents the user with incremental refinements of the final result, according to the solutions and techniques presented in [SASS15], [AS13] and [MPG*14].

3. The TIM Scenario

As a National mobile operator, covering almost 100% of the Italian population and facing strong competition, TIM needs to carefully plan its local marketing campaigns that are based on a number of different factors, sometimes conflicting. Moreover, since these campaigns are expensive, it is possible only to cover few areas and it is very important to make the right choices for the best Return on Investment (ROI). Several parameters have to be considered for the decision making process in order to support quite different high level marketing strategies. As an example, in some cases the managers want to target provinces where the company already have a huge customer base, in order to offer advanced services, while in other cases the managers want to invest in provinces where the company is weak, in order to start a progressive growth process.

The typical decision making process, cooping with 20 regions and 110 provinces, evolves as follows.

1. First the marketing team try to get a global overview of the TIM penetration.
2. Once the global situation is clear, the marketing team is ready to select a number of provinces based on three main parameters: average income, percentage of TIM customers with respect to the total population, number of potential customers in that area.
3. The candidate provinces are subsequently evaluated, focusing on geographical, social, and economical parameters in order to select an optimized subset (max 10 provinces) on which to focus the marketing campaign.
4. This subset is then visualized in a Sankey diagram where the position and direction of each line immediately give the user a clear idea of the marketing and geographical context of each of them, allowing the prioritization or further refinement of the marketing campaign.

4. The Visual Analytics System

The designers involved in the project have used the Systemic Design model [Bis11], paying attention to the organization, optimization and understanding of every single factor at play, focusing at first on user requirements and then highlighting the best conditions and the most interesting facets to work on, while keeping an eye on their mutual relations. That leads to a visualization based on well known visual representations (parallel coordinates and scatter-plot), organized to better explain data and their relationships, without losing in clearness, complexity, organization, appeal, and pleasantness. The forms are adapted to the nature of the information [Cai12] and the data granularity allows for different in-depth levels of analysis, ranging from regions to provinces, and using three different views, described in the following.

The Geographic view represents the penetration of TIM at region level, encoded on a grey scale of color. This is quite intuitive and the decision manager can easily spot regions that present the most space for improvement and profit.

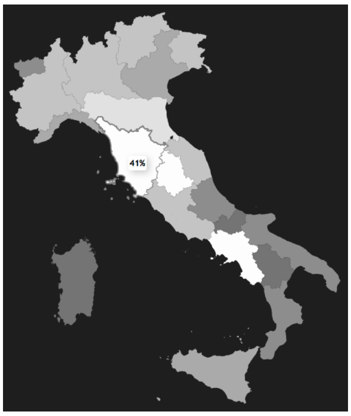

Figure 1: *The thematic map allows for getting an overview of the TIM penetration on region basis.*

The parallel coordinates diagram (see Figure 2) shows the flow among two values: market penetration (percentage of population in possession of a SIM card in the concerned territories) and potential market (difference between the population index and the absolute value of the potential market). The position of axes has a high impact on the patterns emerging from the visualization of samples [HW13]: the external axes represent the data aggregated at the regional level and respectively, from the left to the right, the market penetration and the potential market. The inner axes represent the same values mentioned above, but they are aggregated at the provincial level. All the axes are ordered according to size, from the maximum (top) value to the minimum (bottom). For all four axes, the width of the flow is proportionally sized to the quantity of population of the corresponding regions and provinces. The connection between a region and its belonging provinces is conveyed visually by the use of a sankey diagram: the regional area is proportionally split in the flows, each of them representing the contribution of the corresponding provinces to the overall value. For parallel coordinates diagrams we chose the colors yellow and blue to identify the market penetration: the yellow color represents an above median value, and the blue one represents a below median value. Specifically, we chose this pair of colors to guarantee the right visualization also to the people affected by color-blindness.

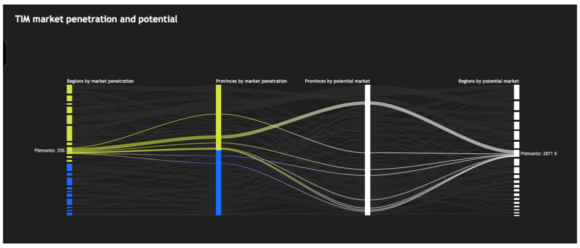

Figure 2: *The interactive parallel plot allows for decomposing regional data on province level, getting an overview of the single region's components.*

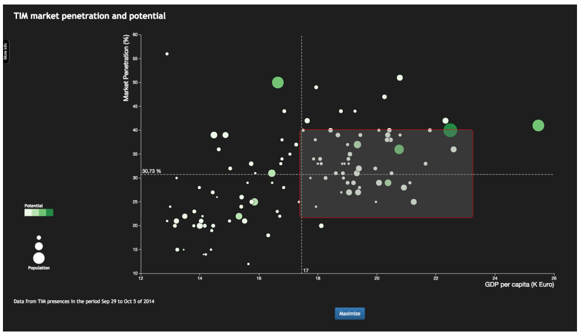

Figure 3: *The scatterplot allows for selecting group of provinces that satisfy some high level campaign scenarios. In the example the analyst has selected provinces characterized by high income and a penetration close to the median; the objective is to promote some additional non basic features (e.g., fast network services) that will likely be accepted by either potential users and TIM customers, in a scenario in which TIM has a good chance of increasing its presence and people GDP is above the median.*

A third visualization is in the form of a scatter-plot diagram; it shows the relationship between the market penetration in the vertical axis and the Gross Domestic Product per capita (GDP) in the horizontal axis. The size of a dot represents the population's amount. The color, ranging from pure white to green represents the potential market. Such a scatterplot is used to start the optimizing algorithm, whose result is presented on the optimized parallel coordinates plot, see Figure 4, highlighting the ten provinces that optimize the result of the objective function.

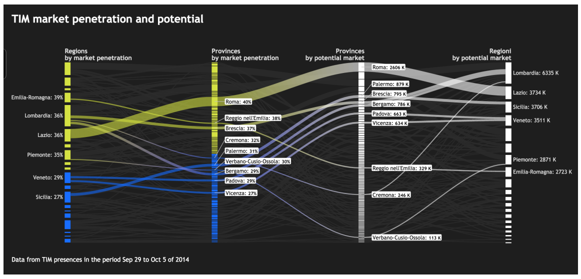

Figure 4: *The result of the optimizing process is presented on the parallel plot and allows the analyst to focus on the most convenient 10 provinces.*

4.1. Interaction

The interactive optimization process is composed of three main steps which define three different views: the parallel coordinates view (Figure 2), the scatter-plot view (Figure 3), and the incremental optimized parallel coordinates view (Figure 4). The exploration starts with the parallel coordinates view which allows the user to explore the data and the relationship between data entities; in the specific, the user can see the fluxes highlighted by using the mouse-over function on them. If there is a flux of interest, she may decide to maintain it highlighted and continue the exploration, with all the other fluxes shown in the background with a dark grey tone, in order to maintain the context. The second step of the analysis is based on the scatter-plot view; this view is particularly useful for spatially grouping together elements with same behaviour (high profit, low penetration, etc.). The user can inspect elements one by one with the mouse and checks details in the associated tool-tip. After making sense of the data, the user is able to choose the set of provinces on which launch the optimization function (by clicking on the "maximize" button). The results of the automatic analysis will be incrementally shown in the incremental optimization view; this view still uses the same visual representation of the parallel coordinates view, in order to maintain visual consistency; this time, however, on the parallel coordinates will be automatically highlighted the provinces that are contained in the result of the objective function, with explicit labels. The user mouse-hovering on them will be able to inspect details regarding the value of the objective function. Every time a better result is produced, this visualization will be updated accordingly. In order to avoid change blindness, the transition between an optimized result and the next is managed in the following way: first the provinces that are no longer in the top ten disappear from the visualization; after that the provinces contained in the new result are inserted one by one. In this way the manager attention is captured by the subset of provinces that remains stable during the optimization steps: additionally, by the use of sequential entering of new provinces, the manager can contextualize each new element with respect to the actual state, capturing all the changes happening to the top ten.

4.2. The Analytical component

The main goal of the Telecom analysts is to select a subset (around 10) of Italian provinces that satisfy two main constraints: a) they must be suited for the actual campaign scenario, like, e.g., rising the penetration percentage of Tim contracts, increasing the profit, etc., and b) they must maximize some marketing goals. The first constraint can be satisfied selecting a subset of the 110 Italian provinces on the scatterplot depicted on figure 3 according to the required general campaign characteristics and detailed in Section 4.1. The typical selection involves 30-50 provinces. The next constraint requires to further restrict the initial subset to select a pilot group of up to ten provinces on which to focus the promotion campaign. However, this selection is not a trivial visual task, according to the heterogeneity of the region-provinces-market penetration-potential market relationships, as shown on Figure 2 . To address this issue, we have designed an Objective Function with Telecom analysts that allows to analytically select the n (top n) most promising provinces of a subsets. The data available for each province p, coming form both public open data data and internal Telecom sources, is:

- *penetration(p)*, i.e., the number of TIM contracts in the observed province (usually expressed as percentage);
- *market(p)*, i.e., the number of people that do not own a TIM contract;
- *gdp(p)*, i.e., the province GDP; and
- *traffic*(p_i, p_j), i.e., the hourly volume of TIM traffic between provinces p_i, p_j.

Discussion with TIM analyst produced the following objective function:

$$\max \sum_{i,j=1}^{n} traffic(p_i, p_j) * \sum_{i=1}^{n} FValue(p_i)$$

where

$$FValue(p) = 0.3 * market(p) + 0.5 * gdp(p) + 0.2 * (1 - |penetration(p) - median(penetration)|)$$

The rationale is that the n pilot provinces should present a strong intra-province traffic together with a balanced combination of potential market, high income, and high likelihood of increasing the penetration (the analysts assumption is that the closer to the median the penetration is, the greater the likelihood of increasing it).

The problem with this objective function is that all values come from (large) tabular relations and that a closed formula for it does not exist, making usual optimization techniques unfeasible for the case at hand. As a consequence, a computation of all the actual tabular values is needed. However, the combinatory explosion of the possible combinations makes this otherwise trivial task unfeasible. As an example, considering the case in which the analyst selects 40 provinces and $n = 10$; that requires to explore a number of combinations equal to the binomial coefficient $\binom{40}{10}$ that is about $8.5 * 10^8$. The processing speed on a high-end desktop is about 10,000 calculations per second. Distributing the computation on ten cores or using a server ten times faster allows for (optimistically) reaching 100,000 calculations per second, getting the exact maximum value in about 2.5 hours. While this is not a hard constraint, it makes an interactive explorative analysis impossible. To solve this problem we implemented our solution according to the emerging technique called progressive or incremental visualization (see, e.g., [SASS15] , [AS13], [MPG*14]), in which we present the user with an initial partial result in a short time and progress with the optimization while the analyst is exploring it. Our solution foresees a burst calculation of 20,000 values of the objective function before presenting the result to the user; this step requires about 4 seconds from selection to visual presentation. After that, while the user is exploring the result, the algorithm proceeds in background, gently updating the visualization each time a better value is computed. If the user is not happy with the result he can quickly start a new exploration, selecting another subset of provinces and getting a new result in 4 seconds. When he is confident that the actual pilot set of top ten provinces is close to his wishes, he can schedule the full optimization in background, getting the optimum later on. The above figures clarify the role of the visualization: even if an objective function is available, its application is not possible on the whole dataset ($\binom{110}{10}$ is about $1.7 * 10^{20}$, requiring $1.7 * 10^{15}$ seconds, $1.3 * 10^9$ years): the scatterplot is used to restrict the input to a suitable subset, making the objective function tractable. However, finding the optimum still requires an incremental visualization that feeds an interactive environment, allowing for visually exploring approximate solutions and quickly discarding non promising ones.

5. Conclusions & Future Work

This paper presented a solution supporting one of the decision making processes of TIM. It encompasses a geographic map and parallel coordinates, allowing the user to get an overview of TIM penetration and market opportunities. An analytical component, fed by the interaction with a scatterplot, provides an optimized list of provinces. In order to insure a fluid interaction, the system presents the user with an initial approximation, refining it along the way. For future development we plan to introduce a number of pre-defined objective functions, each aiming at a different marketing approach (i.e., enter a new region, address the most developed part of the country, improve regions where TIM already has a good market share, etc.), so that an executive can easily select one of them without the need of knowing about math. Another important improvement we are currently working on are visual indicators making the analyst aware of the progress of the converging algorithm and of the confidence level of the current approximation.

References

[AS13] ANGELINI M., SANTUCCI G.: Modeling incremental visualizations. In *Proc. of the EuroVis Workshop on Visual Analytics (EuroVA 2013)* (2013), pp. 13–17. 2, 4

[Bis11] BISTAGNINO L.: Systemic design: designing the productive and environmental sustainability. 2

[Cai12] CAIRO A.: *The Functional Art: An introduction to information graphics and visualization*. New Riders, 2012. 2

[HW13] HEINRICH J., WEISKOPF D.: State of the art of parallel coordinates. *STAR Proceedings of Eurographics 2013* (2013), 95–116. 2

[KBB*10] KEIM D. A., BAK P., BERTINI E., OELKE D., SPRETKE D., ZIEGLER H.: Advanced visual analytics interfaces. In *Proceedings of the International Conference on Advanced Visual Interfaces* (2010), ACM, pp. 3–10. 1

[KBH06] KOSARA R., BENDIX F., HAUSER H.: Parallel sets: Interactive exploration and visual analysis of categorical data. *IEEE Transactions on Visualization and Computer Graphics 12*, 4 (July 2006), 558–568. URL: http://dx.doi.org/10.1109/TVCG.2006.76, doi:10.1109/TVCG.2006.76. 2

[KKEM10] KEIM D. A., KOHLHAMMER J., ELLIS G., MANSMANN F.: *Mastering the information age-solving problems with visual analytics*. Florian Mansmann, 2010. 1

[KMJE12] KO S., MACIEJEWSKI R., JANG Y., EBERT D. S.: Marketanalyzer: an interactive visual analytics system for analyzing competitive advantage using point of sale data. In *Computer Graphics Forum* (2012), vol. 31, Wiley Online Library, pp. 1245–1254. 2

[LT10] LAURSEN G., THORLUND J.: *Business analytics for managers: Taking business intelligence beyond reporting*, vol. 40. John Wiley & Sons, 2010. 1

[MPG*14] MUHLBACHER T., PIRINGER H., GRATZL S., SEDLMAIR M., STREIT M.: Opening the black box: Strategies for increased user involvement in existing algorithm implementations. *Visualization and Computer Graphics, IEEE Transactions on 20*, 12 (2014), 1643–1652. 2, 4

[RSE09] RUDOLPH S., SAVIKHIN A., EBERT D. S.: Finvis: Applied visual analytics for personal financial planning. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on* (2009), IEEE, pp. 195–202. 1

[SASS15] SCHULZ H.-J., ANGELINI M., SANTUCCI G., SCHUMANN H.: An enhanced visualization process model for incremental visualization. *IEEE Transactions on Visualization and Computer Graphics* (2015). 1, 2, 4

[SLFE11] SAVIKHIN A., LAM H. C., FISHER B., EBERT D. S.: An experimental study of financial portfolio selection with visual analytics for decision support. In *System Sciences (HICSS), 2011 44th Hawaii International Conference on* (2011), IEEE, pp. 1–10. 1

[SME08] SAVIKHIN A., MACIEJEWSKI R., EBERT D. S.: Applied visual analytics for economic decision-making. In *Visual Analytics Science and Technology, 2008. VAST'08. IEEE Symposium on* (2008), IEEE, pp. 107–114. 1

[ZNK08] ZIEGLER H., NIETZSCHMANN T., KEIM D. A.: Visual analytics on the financial market: Pixel-based analysis and comparison of long-term investments. In *Information Visualisation, 2008. IV'08. 12th International Conference* (2008), IEEE, pp. 287–295. 2